

# Deutsche Sprache

Zeitschrift für Theorie, Praxis, Dokumentation

Im Auftrag des  
Instituts für deutsche Sprache, Mannheim  
herausgegeben von Hugo Steger (Geschäftsführung),  
Odo Leys, Gerhard Stickel und  
Johannes Schwitalla

15. Jahrgang 1987



ERICH SCHMIDT VERLAG

Herausgeberbeirat: Werner Besch, Bonn; Ulrich Engel, Mannheim; Josef Gerighausen, München; Karl Hyldgaard-Jensen, Kopenhagen; Eijiro Iwasaki, Tokio; János Juhász, Budapest; Gottfried Kolde, Genf; Hans Moser, Innsbruck; Leslie Seiffert, Oxford; Paul Valentin, Paris.

Schriftleitung: Günter Kochendörfer, Freiburg; Ulrich Wetz, Mannheim.

ISSN 0340-9341

© Erich Schmidt Verlag GmbH & Co., Berlin 1987

Satz: C. W. Niemeyer, Hameln

Druck: Poeschel & Schulz-Schomburgk, Eschwege

Nachdruck verboten · Alle Rechte vorbehalten

Die Zeitschrift sowie alle in ihr enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlages. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

## Dokumentation

*Sylvia Dickgießer*

### BERICHT ÜBER DIE 11. JAHRESTAGUNG DER GESELLSCHAFT FÜR LINGUISTISCHE DATENVERARBEITUNG

(Göttingen, 23.–27. Februar 1986)

Die Gesellschaft für Linguistische Datenverarbeitung e.V. (GLDV) ist ein noch junger Fachverband. Sie wurde 1975 unter dem Namen LDV-Fittings gegründet und zählt inzwischen neben 200 Personen auch drei Institute zu ihren Mitgliedern: das Institut für deutsche Sprache, Mannheim, das Max-Planck-Institut für europäische Rechtsgeschichte, Frankfurt, und das Zentrum für Datenverarbeitung der Universität Tübingen. Der in der Vereinssatzung genannte „Zweck der Vereinigung ist die Förderung der wissenschaftlichen linguistischen Datenverarbeitung mit dem Ziel der Kooperation und Koordination von Forschungsvorhaben.“ Mit dem ein- bis zweimal jährlich erscheinenden „LDV-Forum“ verfügt der Verein über ein Publikationsorgan, das sich bei anhaltendem Engagement der ehrenamtlich tätigen Redaktion in den nächsten Jahren durchaus zur Fachzeitschrift entwickeln könnte.

Zur 11. Jahrestagung der GLDV mit dem Thema „Linguistische Datenverarbeitung in den Geisteswissenschaften“ trafen sich zirka 150 Teilnehmer im geisteswissenschaftlichen Zentrum der Universität Göttingen.

‘Linguistische Datenverarbeitung’ ist eine umstrittene Bezeichnung, die für maschinelle Sprachdatenverarbeitung zu linguistischen Zwecken ebenso verwendet wird wie z. B. im Namen des Vereins, als Sammelbezeichnung für Computerlinguistik, Informationslinguistik, Informationswissenschaft, maschinelle Sprachverarbeitung etc.

In diesem weitesten Sinne stand Linguistische Datenverarbeitung in Göttingen auf dem Programm, das neben dem einführenden Referat von **Mart Rimmel** (Tallinn) über „Die mathematischen Mittel für die Formalisierung der historischen Linguistik“ etwa dreißig Vorträge in acht Sektionen umfaßte:

- Informationslinguistik
- Philosophie und Logik von Frage-Antwort-System
- Formale Grammatiken
- Geisteswissenschaftliche Analysesysteme
- Automatische Analyse und Synthese natürlicher Sprache
- Maschinelle Lexikographie
- Maschinelle Übersetzung
- Textdatenverarbeitung

Leider standen auch in Göttingen nur dreißig Minuten für jeden Vortrag (einschließlich der anschließenden Diskussion) zur Verfügung. Infolge der knappen Zeit waren einige Darstellungen notwendigerweise sehr komprimiert, was nicht auf dem gleichen Gebiet arbeitenden Zuhörern einen Zugang möglicherweise erschwerte. Bedauerlich finde ich dies vor allem angesichts der Notwendigkeit, Experten so entfernter Gebiete wie Datenverarbeitung und Geisteswissenschaften einander näherzubringen, um effektive und sinnvolle Instrumente Linguistischer Datenverarbeitung entwickeln zu können.

In zwei Podiumsdiskussionen wurden die Themen „Computerlinguistik und philologische Datenverarbeitung. Kontroversen und Perspektiven“ und „Brauchen Lexikographen wirklich die Werkzeuge der Linguistischen Datenverarbeitung?“ behandelt.

Besonders erwähnenswert sind die intensiven Bemühungen der GLDV um Ausbildungsmöglichkeiten und ein entsprechendes Informationsangebot an Studenten. Am Nachmittag des zweiten Konferenztages spiegelte sich diese schon traditionelle Studentenfremdlichkeit in einem dreistündigen „Informationsmarkt LDV-Studium“, organisiert von **Magdalene Lutz-Hensel**, die auch für den „Studienführer Linguistische Datenverarbeitung“ verantwortlich zeichnet. Der Informationsmarkt mußte nicht mit Vorträgen konkurrieren. Außer der anschließenden Mitgliederversammlung fanden an diesem Nachmittag keine weiteren Veranstaltungen statt.

Am Nachmittag des 27. 2. tagten die Arbeitskreise der GLDV: „Studiengänge mit LDV-Anteil“, „LDV und Nachbarn“, „Mikros und Textdatenverarbeitung“, „Sozialwissenschaften und LDV“, „Maschinelle Übersetzung“, „Maschinelle Lexikographie“, „Textanalyse“.

Im folgenden möchte ich einen kurzen Überblick über die von mir besuchten Veranstaltungen geben.

### **Informationslinguistik**

Über Arbeiten am Projekt PADOK in Regensburg berichteten **Christine Schneider** und **Christa Womser-Hacker** in ihrem Vortrag „PADOK. Test und Vergleich von Texterschließungssystemen für das Deutsche Patent- und Fachinformationssystem“ sowie **Jürgen Krause** in einem Vortrag über „Informationslinguistische Komponenten von Informationssystemen. Konsequenzen aus den Ergebnissen der ALP-Studie und PADOK“. Im Projekt PADOK sollen Entscheidungshilfen für ein Patentedokumentations- und Informationssystem erarbeitet werden.

Schneider und Womser-Hacker stellten Ergebnisse eines Vergleichs von Texterschließungsleistungen der Systeme FREITEXT, PASSAT, CTX und DETECT vor, die mit einer Testdatenbank erzielt wurden. Es handelte sich um Daten des Deutschen Patentamtes, die im Vergleich z. B. zu Lexikonartikeln simpel strukturiert sind: Titel, Abstract (ca. 120 lfd. Wörter).

Es war interessant zu hören, daß die Retrievalergebnisse anscheinend in hohem Maße von speziellen Datenbankerfahrungen der Benutzer (Patentprüfer) abhängig waren. Prüfer, die vor dem Test bereits mit anderen als den Testsystemen PASSAT und CTX gearbeitet hatten, erzielten mit den Testsystemen schlechtere Ergebnisse als unerfahrene Benutzer.

Krause wies auf die sprachlichen Spezifika der Testdaten hin (hohes Abstraktionsniveau, komplexe Nominalphrasen, wenige Wortneubildungen ect.) und stellte einen Zusammenhang her zwischen sprachlichen Besonderheiten, Systemkomponenten und Systemleistungen.

## Geisteswissenschaftliche Analysesysteme

**Karin Dierks** und **Christian Zinko** (Universität Graz) berichteten über „Computerunterstützte Emendierung des Hethitischen“. Im Mittelpunkt der philologischen Arbeiten stehen Graphemanalysen, Ergänzungen unvollständig überlieferter Wörter, Erstellung vor thesaurusähnlichen Wörterbüchern, Rekonstruktion fragmentarischer Texte und die grammatische Beschreibung des Hethitischen. Neben thesaurusähnlichen Wörterbüchern aus Texten werden mit Computerhilfe Belegstellenverzeichnisse und KWIC-Konkordanzen erstellt. Entsprechende Emendierungsalgorithmen ermöglichen einen Computereinsatz auch bei der Wort- und Textergänzung. Den Mitarbeitern steht leider nur ein minimales Kontingent an Rechenzeit für den Dialogbetrieb zur Verfügung. Einsparungen dieser Art sind insofern fragwürdig, als viel intellektuelle Kapazität in die Suche nach Lösungen dadurch entstehender Probleme investiert werden muß.

Über Probleme bei Erhebung, Kodierung und Retrieval von Sprachdaten für den Südwestdeutschen Sprachatlas referierte **Bernhard Kelle** (Universität Freiburg). Sein Thema lautete: „Retrieval in mundartlichem Sprachmaterial ohne feste Wortgrenzen“. Die übliche Methode, große Bestände maschinenlesbarer Sprachdaten durch automatische Lemmatisierung so zu erschließen, daß mit Hilfe von Suchprogrammen schnelles und sicheres Recherchieren möglich wird, scheitert an den Besonderheiten des variantenreichen dialektalen Materials. Eine manuelle Lemmatisierung scheint angesichts der Zahl von zwei Millionen Belegen als zu aufwendig. Durch Reduktion der Belege auf ihre Konsonantenstruktur hofft man in Freiburg, den Variantenreichtum bewältigen und den Einsatz automatischer Suchverfahren ermöglichen zu können.

Die Teilnehmer der Podiumsdiskussion über „Computerlinguistik und philologische Datenverarbeitung“ repräsentierten ein breites Interessenspektrum, das ich gerne, wie sie es taten, mit Hilfe einiger (ihrer) Stichwörter vorstellen möchte:

**Gärtner** (Trier): Philologische Datenverarbeitung;

**Hoepfner** (Hamburg): Künstliche Intelligenz (natürlich-sprachliche Schnittstellen);

**Ludwig** (Tübingen): Literaturwissenschaftliche Stilanalyse, Literaturkritik;

**Lutz** (Koblenz): Angewandte Informatik, Linguistik, Künstliche Intelligenz (natürlich-sprachliche Schnittstellen);

**Ott** (Tübingen): Literarische und Dokumentarische Datenverarbeitung, Editionstechnik, Texterschließung, Bibliographie;

**Schweißthal** (München): Phonetik, Dialektologie, Sprachdatentechnik;

**Hellwig** (Heidelberg): Textlinguistik, Computerlinguistik, Syntax, Semantik, Künstliche Intelligenz (natürlich-sprachliche Systeme).

In der Diskussion trat die Disparität der Benennungen bzw. die Heterogenität dessen, was 'Computerlinguistik' und 'philologische Datenverarbeitung' genannt wurde, deutlich zutage. Die Sprachverwirrung innerhalb und/oder zwischen den Disziplinen wurde dann auch Diskussionsthema. Dieser Problemkomplex zeigte sich neben einem anderen, der von Lutz thematisiert, für das Rahmenthema der Tagung charakteristisch sein dürfte: Die Umsetzung qualitativer Begriffe der Geisteswissenschaften in 'komparative' Begriffe der DV.

## Maschinelle Lexikographie

In der von **Drewek** (Zürich) geleiteten Podiumsdiskussion zum Thema „Brauchen Lexikographen die Werkzeuge der Linguistischen Datenverarbeitung?“ trafen DV-Experten Gesprächspartner mit auffallend unterschiedlichen Erfahrungen mit Mitteln der Datenverarbeitung.

**Birkenhauer** (Europäisches Übersetzerkollegium, Straelen) berichtete über die Übersetzer-tätigkeiten am Kollegium und daraus resultierende kleine Wörterbuchprojekte. Bisher wird in Straelen lediglich mit Textverarbeitungssystemen auf Mikrocomputern gearbeitet. Die verwendeten Systeme sind komfortable Werkzeuge für die Texterstellung und -editierung, bieten jedoch oft nur beschränkte Möglichkeiten, wenn es darum geht, das mit ihrer Hilfe erfaßte Wortmaterial während des Übersetzungsprozesses zur Verfügung zu stellen.

**Schölze-Stubenrecht** (Duden-Redaktion, Mannheim) erwähnte, daß auch in der Duden-Redaktion bisher lediglich Erfahrungen mit Textverarbeitungssystemen gesammelt werden konnten. Wie Birkenhauer erwartete auch er von der Tagung u. a. Informationen über adäquate DV-Werkzeuge.

**Hellmann** (Institut für deutsche Sprache, Mannheim) skizzierte (s)einen lexikographischen Arbeitsplatz am IdS und die Unterstützung der Arbeiten am Maschinellen Korpuswörterbuch (MKWB) durch die Bereitstellung maschinenlesbarer Daten und Auswertungsprogramme.

**Frackenpohl** (Köln) gab Hinweise auf bestehende DV-Werkzeuge insbesondere für Mikrocomputer und deren Brauchbarkeit für die lexikographische Arbeit.

**Storrer** (Wissenschaftliches Zentrum der IBM, Heidelberg) berichtete über ihre Arbeit an der Wörterbuchkomponente von LEX, eines juristischen Expertensystems.

Die Diskussion zeigte die Notwendigkeit eines intensiveren Kontaktes zwischen Lexikographen und DV-Experten (Computerlinguisten, ...), intensiveren Austausches von Daten und Programmen und größerer Aufmerksamkeit für Fragen der Anlagenunabhängigkeit (Portabilität, Kompatibilität) von Daten und Software. Wieder einmal wurde deutlich, daß 'maschinelle Lexikographie' zu differenzieren ist in 'computerunterstützte Lexikographie' und 'Lexikographie zur Computerunterstützung'.

Im Anschluß an die Podiumsdiskussion sprach **Andreas Blumenthal** (Universität Heidelberg) zum Thema „Vorüberlegungen zur Möglichkeit des Computereinsatzes bei der Abfassung lexikographischer Bedeutungserläuterungen. Ein Bericht aus dem Projekt COLEX“. COLEX ist ein Akronym für 'Computerunterstützte Lexikographie'. Im wesentlichen sagte Blumenthal: Voraussetzung eines Computereinsatzes in der Lexikographie, der über die Bereitstellung von Quellen und Unterstützung in der Edition hinausreicht, ist eine Explikation lexikographischer Tätigkeiten als wohldefinierte Operationen auf Daten mit bestimmten (wohldefinierten) Strukturen. Seiner Ansicht nach wurden in den Gebieten Künstliche Intelligenz und Information und Dokumentation Systemkomponenten und Techniken erarbeitet, die auf ihre Brauchbarkeit für lexikographische Zwecke zu prüfen sind.

**Gert Frackenpohl** referierte über den „Aufbau einer Lexikondatenbank des Vokabulari Rumantsch Grischun“, insbesondere über Probleme der Informations- und Datenstruk-

turierung in einem zweisprachigen Wörterbuch, das mittels eines kommerziellen relationalen Datenbanksystems gespeichert wird. Die Datenbank wurde im Auftrag der Lia Rumantscha in Chur erstellt. Zur Zeit wird sie noch auf einem PC betrieben. Der parallele Betrieb auf einem Großrechner ist geplant.

In seinen „Reflexionen zum Begriff Wortdatenbank in der linguistischen Datenverarbeitung“ machte **Hans Huonker** (Zürich) deutlich, wie unterschiedlich Informationsklassen, Strukturen und Zwecke bestehender Wortdatenbanken des Deutschen sind. Er skizzierte ein semantisches Kategoriensystem (semantische Tiefenkasus), das eine Integration der unterschiedlichen semantischen Informationsklassen dieser Datenbanken ermöglichen soll.

Sitzung des Arbeitskreises „Maschinelle Lexikographie“:

Dieser AK gehört zu den aktivsten und expansivsten Arbeitskreisen der GLDV. Wissenschaftliche Zentren der IBM sind darin ebenso repräsentiert wie z. B. die Universitäten Koblenz und Heidelberg, das Institut für Kommunikationsforschung und Phonetik der Universität Bonn (IKP), das Institut für deutsche Sprache in Mannheim und die Gesellschaft für Information und Dokumentation in Frankfurt. Zur Zeit trifft der Arbeitskreis Vorbereitungen für ein Tutorium, das im Rahmen der Jahrestagung 1987 veranstaltet werden soll. Zur Zielgruppe des Tutoriums gehören Studenten, Lexikographen ohne bzw. mit geringen DV-Kenntnissen, Experten aus dem DV-Bereich ohne bzw. mit geringen lexikographischen Kenntnissen.

Mindestens drei große Themenbereiche stehen auf dem Programm:

- Lexikographische Arbeit – Lexika
- Einführung in die DV
- Bereitstellung maschinenlesbarer Texte und automatischer Texterschließungsverfahren

Dieser Plan verdeutlicht das Bemühen von GLDV-Mitgliedern, Voraussetzungen interdisziplinärer Verständigung zu schaffen und zu verbessern und verdient regen Zuspruch, denn ohne diese Basis ist linguistische Datenverarbeitung in engem Bezug zu den Erfordernissen sprachwissenschaftlicher Arbeit nicht möglich.

### Textdatenverarbeitung

**Karin Haenelt** (Universität Heidelberg) stellte das Projekt PRO TEXT vor. Ihr Vortragsthema lautete: „Ein sprachtheoretisches Modell geisteswissenschaftlicher Texterschließungsverfahren als Grundlage einer Softwareentwicklung.“ Das Akronym PRO TEXT steht für 'Programmpaket mit endbenutzerfreundlichem Bedienungsrahmen zur Unterstützung geisteswissenschaftlicher Texterschließung in einem engen Verbund von Arbeitsplatzrechnern und Zentralrechnern'. Wie in COLEX (Pilotanwender von PRO TEXT) geht es in PRO TEXT zunächst darum, ein Modell linguistischer und philologischer Texterschließung zu erarbeiten. Auf der Grundlage dieses Modells soll beurteilt werden, welche Tätigkeiten formalisierbar und damit automatisierbar sind. Wie Blumenthal betont Haenelt, daß bei der Bestimmung der Grenzen des Computereinsatzes sogenannte fortgeschrittene Techniken der Computerlinguistik und KI-Forschung zu berücksichtigen sind.

Haenelts Beispiele für mögliche Anwendungen machten deutlich, daß die Definition geisteswissenschaftlich relevanter Phänomene keine geringen Anforderungen stellen

wird. Vermutlich kommt hier das von Lutz thematisierte Problem der Operationalisierung qualitativer Begriffe zum Tragen. Es dürfte interessant sein zu beobachten, ob und wie das von der DV bereitgestellte Instrumentarium geisteswissenschaftliche Erkenntnisinteressen und Ergebnisse beeinflusst.

### **Maschinelle Übersetzung (MÜ)**

Sämtliche Vorträge dieser Sektion wurden von Mitarbeitern Saarbrücker Forschungsprojekte bestritten. Ich hörte **Erich Steiner**: „Die Zuweisung satzsemantischer Rollen im Rahmen des MÜ-Projektes EUROTRA-D“ und Cornelia **Zelinsky-Wibbelt**: „Entwurf eines Systems semantischer Merkmale für die maschinelle Übersetzung“.

Steiner stellte ein auf der Systemic Functional Grammar basierendes Verfahren vor, mit dessen Hilfe Verben, die in einer Lexikonkomponente enthalten sein sollen, Kasusrahmen zugeordnet werden. Die manuell erstellte Lexikonkomponente dient einer automatischen Syntaxanalyse von Texten. Das Verfahren soll v. a. eine konsistente semantische Beschreibung ermöglichen.

Zelinsky-Wibbelt entwarf ein Basissystem semantischer Merkmale, das, nach gestaltpsychologischen Prinzipien strukturiert, eine Kategorisierung menschlicher Erfahrung ermöglichen soll. Ziele des Entwurfs sind eine interlinguale Kompatibilität der Beschreibungskategorien und Sicherung der Konsistenz von Beschreibungsergebnissen.

Die Diskussion beider Vorträge verdeutlichte, in wie hohem Maße die Konzeption von Kategoriensystemen in diesem Arbeitsbereich von pragmatischen Gesichtspunkten bestimmt wird. Hinweise auf mangelnde Beschreibungsadäquatheit der Konzepte wurden bestätigt. Ihnen wurden jedoch Hinweise auf gute Operationalisierbarkeit und Brauchbarkeit der Konzepte im Hinblick auf die o. g. Projektziele gegenübergestellt.

### **Resümee**

Dank der Vorbereitungen des Organisationskomitees, für das **Ursula Klenk**, Romanistisches Seminar der Universität Göttingen, **Manfred Thaller**, Max-Planck-Institut für Geschichte, und **Peter Scherber**, Gesellschaft für Wissenschaftliche Datenverarbeitung, stellvertretend genannt seien, nahm die Tagung einen reibungslosen Verlauf. Alle von mir besuchten Veranstaltungen hatten einen sehr positiven Aspekt gemeinsam: eine Atmosphäre, die es erlaubte, über Probleme zu sprechen. Die Teilnahme vieler Wissenschaftler mit Arbeitsschwerpunkten außerhalb der DV förderte die Diskussion konkreter Anforderungen an DV-Entwicklungen. Vorführungen von Hard- und Software in den Tagungspausen veranschaulichten einige neue Entwicklungen, womit man dem Informationsbedürfnis der in Datenverarbeitung wenig erfahrenen Teilnehmern entgegenkam.

Daß DV-Experten die Vagheit und Disparatheit vieler ihrer Benennungen in dieser Deutlichkeit problematisierten, betrachte ich als einen sehr nützlichen Aspekt dieses interdisziplinären Treffens, v. a. angesichts der von Lutz akzentuierten negativen Folgen bestehender Inkonsistenzen für die Ausbildung.



### **Ausblick**

Die Jahrestagung 1987 ist in Zusammenarbeit mit der Deutschen Gesellschaft für Sprachwissenschaft in Darmstadt geplant. Thema: „Computermodellierung linguistischer Theorien“. Sollte die Zusammenarbeit mit der DGfS nicht zustande kommen, gibt es zwei Alternativen. Die Tagung findet dann entweder am IKP in Bonn zum Thema „Spracherkennung – Sprachsynthese“ statt oder an der Universität Saarbrücken zum Thema „Maschinelle Übersetzung“.

Sylvia Dickgießer

Institut für deutsche Sprache, Friedrich-Karl-Str. 12, D-6800 Mannheim